

Non-stationary patterns of isolation-by-distance: inferring measures of genetic friction.

Nicolas Duforet-Frebourg¹, Michael G.B. Blum¹

¹ Université Joseph Fourier, Centre National de la Recherche Scientifique, Laboratoire TIMC-IMAG, Grenoble, France.

Running Head: Inference of genetic friction.

Keywords: landscape genetics, gene flow, genetic barrier, isolation by distance, non-stationary kriging

Corresponding author: Michael Blum

Laboratoire TIMC-IMAG, Faculté de Médecine, 38706 La Tronche, France

Phone +33 4 56 52 00 65

Fax +33 4 56 52 00 55

Email: michael.blum@imag.fr

Abstract

The pattern of isolation-by-distance arises when population differentiation increases with increasing geographic distances. This pattern is usually caused by local spatial dispersal which explains why differences of allele frequencies between populations accumulate with distance. However, the pattern of isolation-by-distance can mask complex variations of demographic parameters. Spatial variations of demographic parameters such as migration rate or population density generate non-stationary patterns of isolation-by-distance where the rate at which genetic differentiation accumulates varies across space. Barriers to gene flow are particularly well studied examples that generate non-stationary patterns of isolation-by-distance. Using the concept of genetic friction, we develop a statistical method that characterizes non-stationary patterns of isolation-by-distance. Genetic friction at a sampled site corresponds to the local genetic differentiation between the sampled population and fictive populations living in the neighborhood of the sampling site. To avoid defining populations in advance, the method can also be applied at the scale of individuals. The inference of genetic friction relies on a matrix of pairwise similarity between populations or individuals. Typical matrices of similarity include correlation matrices or F_{ST} -based matrices. The proposed framework is appropriate for dealing with massive data because it relies on a pairwise similarity matrix, which can be obtained with computationally efficient methods. A simulation study shows that maps of genetic friction can detect barriers to gene flow but also other patterns such as continuous variations of gene flow across habitat. The potential of the method is illustrated with 2 data sets: genome-wide SNP data for the human Swedish populations, and AFLP markers for alpine plant species. The software *FRICITION* implementing the method is available at <http://membres-timc.imag.fr/Michael.Blum/Friction.html>.

INTRODUCTION

Characterizing the pattern of genetic differentiation within a species is a recurring task in population genetics analyses. WRIGHT (1943) introduced the model of isolation by distance (IBD) which assumes that differences of allele frequencies between populations accumulate under the assumption of local spatial dispersal. Because of local dispersal, IBD models predict the so-called *pattern of IBD* where population differentiation increases with increasing geographic distances (SLATKIN 1993; ROUSSET 1997). This pattern is observed in many model and non-model organisms as well as in humans suggesting that local dispersal is a leading evolutionary force (SHARBEL *et al.* 2000; RAMACHANDRAN *et al.* 2005; HARDY *et al.* 2006; HELLBERG 2009).

However, the pattern of IBD can mask complex variations of demographic parameters resulting in differential increases of genetic differentiation in different regions of the habitat. Variations of demographic parameters can arise when population densities or migration rates vary across space (SLATKIN 1985). With the advent of landscape ecology (PICKETT and CADENASSO 1995) and landscape genetics (MANEL *et al.* 2003), the spatial variation of demographic parameters is an important topic because spatial heterogeneity (or landscape characteristics) is now recognized to be a key factor to explain population differentiation and gene flow (MCRAE and BEIER 2007). Examples of spatial heterogeneity influencing population differentiation include fragmented landscapes in urban and agricultural area where there are ‘corridors’ for gene flow (ARNAUD 2003; MUNSHI-SOUTH 2012). These habitats can also be characterized by spatially varying local subpopulation size, another factor causing spatial variation of the rate of differentiation between populations (SERROUYA *et al.* 2012). Barriers to gene flow, which can be caused by anthropogenic or geographic factors, are also emblematic examples of spatial heterogeneity influencing population structure (e.g. CASTELLA *et al.* 2000; EPPS *et al.* 2005; RILEY *et al.* 2006; GAUFFRE *et al.* 2008; ZALEWSKI *et al.* 2009). Because the identification of the barriers to gene flow has attracted considerable attention (STORFER *et al.* 2010), there is a large variety of statistical methods dedicated to their detection (BARBUJANI *et al.* 1989; BOCQUET-APPEL and BACRO 1994; DUPANLOUP *et al.* 2002; MANNI *et al.* 2004; CERCUEIL *et al.* 2007; CRIDA and MANEL 2007; MANEL *et al.* 2007; SAFNER *et al.* 2011). Here, we propose a more general method that characterizes *non-stationary* patterns of isolation-by-distance. A non-stationary pattern of isolation-by-distance occurs when the spatial rate of differentiation between individuals or between populations depends on space. Non-stationary patterns of IBD arise when there is a barrier to gene flow because there is a larger rate of genetic differentiation in the region of the barrier but also on different situations such as continuous variations of gene flow across habitat.

To characterize non-stationary patterns of IBD, our approach provides a measure of local differentiation at each location where genetic data is available. By analogy with the concept of *friction map* in ecology (RAY 2005), which measures the cost of movement through the landscape, the measure of local differentiation can be thought of as a measure of *genetic friction* with larger friction values indicating larger population genetic differentiation per unit of

spatial distance. The principle of the method is to estimate for each sampled location z_i , $i = 1, \dots, n$, a pairwise measure of population differentiation or of dissimilarity between the population located at the sampled location and fictive populations located in the neighborhood of z_i (see Figure 1). The method is not restricted to pairwise population measurements and can also accommodate individual pairwise measures. Working at the scale of individuals is a desirable feature since using individuals as the operational unit avoids potential bias in identifying populations in advance and offers the opportunity to conduct studies at a finer scale (MANEL *et al.* 2003, 2007). Using a detailed simulation study, we demonstrate that the method can correctly infer local variation of genetic differentiation and we present applications of the method to genome-wide SNP data for the human Swedish population (HUMPHREYS *et al.* 2011) and AFLP makers from alpine plants (GUGERLI *et al.* 2008).

METHODS

We assume that the data consist of the correlation matrix of allele frequencies over populations. To assess local genetic differentiation around a given sampled site, we estimate the correlation of allele frequencies between the sampled population and the neighboring populations that have not been sampled. Neighboring populations are located at a fixed and short distance from the sampled populations, and we measure the expected correlation (averaged over neighbors) of allele frequencies between the sampled population and the neighboring fictive populations. Since we aim at providing friction values that should be larger in regions of abrupt genetic changes, we consider one minus these *local* correlations as friction measures.

We estimate the local correlation using kriging, which refers to geostatistical techniques to interpolate the value of a variable at unsampled locations from observations of its value at nearby and sampled locations (CRESSIE 1993). We do not use kriging in a standard manner, which would consist of interpolating the allele frequencies at the neighboring sites where the fictive populations are located. We instead provide an estimation of the correlation of a variable (allele frequency here) between the sampled sites and the neighboring sites. The kriging procedure is described below.

The kriging approach: A viewpoint of kriging is to model the joint values of the variable at sampled and unsampled sites as a multivariate Gaussian variable (BISHOP 2006)

$$(X, Y) \rightsquigarrow \mathcal{N}(m, \Psi), \quad (1)$$

where

$$\Psi = \begin{pmatrix} \Psi_{xx} & \Psi_{xy} \\ \Psi_{xy} & \Psi_{yy} \end{pmatrix},$$

where X (resp. Y) is the vector of values at the sampled (resp. unsampled) sites, m is a constant mean over the habitat

(ordinary kriging), Ψ_{xx} (resp. Ψ_{yy}) is the covariance matrix of X (resp. Y) and Ψ_{xy} contain the covariances between the elements of X and Y . The interpolation of the variable at the unsampled sites is obtained using the conditional distribution of Y given X , which can be written in the following regression form

$$Y - m = \tau(X - m) + \epsilon, \quad (2)$$

where $\tau = \Psi_{xy} \Psi_{xx}^{-1}$ and ϵ is a residual independent of X (LE and ZIDEK 1992). Using the regression equation (2) that relates the variable values at sampled and unsampled sites, we can show that the predictive value of the covariance matrix between X and Y is given by

$$\Sigma_{xy} = \tau \Sigma_{xx}. \quad (3)$$

where Σ_{xx} is the covariance matrix for the sampled sites and we consider the empirical covariance matrix S in our computations. Providing the correlation instead of the covariance between sampled and unsampled sites requires the standardization of the covariance equation (3) and the renormalization formula is provided in Appendix A. The predictive distribution of the matrix Σ_{xy} is finally obtained by integrating over the posterior distribution of τ , which is a function of the matrix Ψ . The parametric model for Ψ and the sampling algorithm that provides its posterior distribution is given below.

A model for the correlogram:

We consider the standard model of *stationary* kriging which assumes that the correlation between two points only depend on the distance between these two points. Using these assumptions, we should model how the correlation decreases with increasing distance. We assume that this function C , called the correlogram, decays exponentially

$$C(d) = (\alpha + (1 - \alpha)e^{-d/r} + \lambda \mathbb{1}_{d=0}) / (1 + \lambda), \quad (4)$$

where d is the distance between two points, $\mathbb{1}$ denotes the indicator function, α determines the sill, which measures the limiting value of the correlation function, r is the *range* parameter and λ is the *regularization* parameter. The parameter λ is introduced to account for the uncertainty of measure at the sampled sites (BISHOP 2006). More importantly, it ensures that the matrix Ψ_{xx} is invertible, which is required for computing the regression matrix τ . The range parameter r is inversely related to the rate at which correlation decays with distance. We sample (α, λ, r) from the posterior distribution using a Gibbs sampling algorithm with details provided in Appendix B (HANDCOCK and STEIN 1993). It might appear paradoxical to consider the model of stationary kriging to provide a spatial estimation of the different rates at which correlation decays with distance. However, the stationary model of equation (4) only provides values for the regression matrix τ and it is the data contained in Σ_{xx} that convey information about non-stationarity through equation (3).

SIMULATION STUDY

In the simulation study, we consider two different models for generating non-stationary patterns of isolation-by-distance. First, we consider non-homogeneous stepping stone models in one and two dimensions. We simulate with *ms* (HUDSON 2002) spatially-dependent effective migration rate $4N_0m$ where N_0 is the population size of each deme and m is the migration rate per generation between two neighboring demes. The second model is analytic and has been developed for performing non-stationary kriging when the correlogram function (equation(4)) is assumed to vary across space (PACIOREK and SCHERVISH 2006). The range parameter r of equation (4), which measures the rate at which correlation decays with distance, is assumed to be a function of space. For the second model, zone of abrupt changes such as genetic barriers correspond to regions with a smaller range parameter because correlation decays more rapidly with distance in these regions.

Barrier in one-dimensional habitat: We provide an example of a one dimensional barrier to show that the parameters of the correlogram function (equation (4)) affect the estimated values of genetic friction. We simulate a stepping stone model with 100 populations of effective sizes $N_0 = 1000$ diploid individuals. We sample 20 equidistant populations and we consider 20 chromosomes in each of them. Migrations are constant between neighboring populations and $4N_0m = 20$. The barrier is located between populations 50 and 51 and arose 8 units of time ago ($4N_0m = 0$) where time is counted in units of $4N_0$ generations. As input data (Σ_{xx} with $n = 10$), we consider the pairwise correlation of allele frequencies for the 20 sampled populations.

Figure 2 displays the estimated friction values along the one-dimensional habitat when sampling 20 equally-spaced populations. For each sampled deme, the friction corresponds to one minus the expected correlation between the sampled deme and its two neighbors. Whatever the values of the correlogram parameters, (α, λ, r) , the friction values are larger in the middle of the habitat, which is consistent with the presence of a barrier to gene flow. Nonetheless the detailed trajectory of the friction curve depends on the choice of the parameter value. To account for the uncertainty associated with the parameters of the correlogram function, we integrate the friction values over the posterior distribution of the correlogram model (thick line in Figure 2). When considering a random and uniform sampling of the 20 populations instead of a regular sampling, we also find that the friction values are larger around the barrier to gene flow (Figure S1).

Barriers in two-dimensional habitat: We consider two examples of a 2-dimensional habitat with genetic barriers. In the first example of a 2-dimensional habitat, the data are simulated using a stepping-stone model with a 10×10 grid. In each population, there are $N_0 = 1000$ diploids individuals per population and we sample all populations considering 20 chromosomes in each of them. The migration rate between neighboring populations is of $4N_0m = 20$ where migration only occurs along horizontal and vertical lines but not along diagonals. We then assume that two barriers arose $T_1 = 5$ and $T_2 = 3$ units of time ago where time is counted in units of $4N_0$ generations (see Figure 3). In

the second 2-dimensional example, we specify explicitly the local decay of correlation using the non-stationary model of PACIOREK and SCHERVISH (2006). We assume that there are three genetic barriers, which correspond to three different regions where the range parameter (r in equation(4)) is smaller (Figure 4). In both examples, the input matrix of pairwise values (Σ_{xx}) is the correlation matrix for the sampled sites, although the correlation is estimated using the simulated allele frequencies for the stepping stone example and it is obtained analytically using the convolution formula of PACIOREK and SCHERVISH (2006) for the second example.

Figures 3 and 4 show that the friction values are larger around the genetic barriers as expected. For both examples, the relative importance of the barriers is retrieved. The strongest barrier has the largest friction value and the weakest barrier has the smallest friction values. For computing friction values, we also consider F_{ST} pairwise values instead of correlation values in the stepping-stone model. The matrix of input data Σ_{xx} contains the pairwise $(1 - F_{ST})$ values that decay with increasing geographical distance as assumed by the correlogram equation (4). The friction values now correspond to the expected F_{ST} between the sampled populations and their fictive neighbors. We find that the friction map obtained with the F_{ST} measures is similar to the friction map obtained with the pairwise correlation (Figure S2).

A gradient of gene flow: We also consider a different pattern of non-stationary IBD consisting of a 2 dimensional stepping-stone model with a spatial gradient of gene flow. We assume that gene flow is maximum at the lower left corner of the habitat and decreases proportionally with distance from the lower left corner of the habitat (Figure 5). As expected, we find that genetic friction increases when moving away from the lower left corner of the habitat (Figure 5).

This example is particularly interesting instance of non stationarity, which can not be described with barriers to gene flow. For the previous examples, the software *barrier*, which detects zones of abrupt genetic change (MANNI *et al.* 2004), was able to retrieve the barriers in both the one and two-dimensional habitat (Figure S3). However, for the gradient of gene flow, *barrier* incorrectly finds a barrier in the upper right corner of the habitat, which is nonetheless consistent with the fact that gene flow is minimal here (Figure 5). Additionally, multidimensional scaling—another commonly used method to represent differentiation between populations—provides a meaningful representation for the examples of barriers because the populations that live on different sides of the barriers are clearly separated in the two-dimensional representation of MDS (Figure S4). However, the interpretation of the pattern obtained with MDS is much more difficult for the example of a gradient of gene flow (Figure 5). The observed pattern found with MDS is consistent with the gradient of gene flow because populations living in regions of high gene flow (dark points in Figure 5) are located more closely on the MDS plot than populations living in regions of low gene flow (clear points in Figure 5). Although consistent with a gradient of gene flow, the MDS plot is not as easily interpretable as the map of genetic friction for this example.

APPLICATIONS

Non-stationary patterns of IBD among the Swedish population: We first illustrate the kriging methodology using a human SNP data set with a particularly dense geographic sampling. The data consist of genome-wide SNPs for 5174 Swedish individuals that cover all of the 21 Swedish counties (HUMPHREYS *et al.* 2011). To assign each individual to a county, HUMPHREYS *et al.* (2011) used available geographic information with the following order of priority: city or village of birth, county of birth, municipality or city of residence and county of residence if it is the only information available. They found strong differences between the far northern counties and the remaining counties, and additionally demonstrated that the northern counties are more clearly genetically differentiated from each other than the southern counties are from each other.

Since our framework is an extension of isolation-by-distance, we first check that population differentiation increases with increasing geographical distance. We confirm the prevalence of isolation by distance in Sweden ($P < 10^{-7}$ for a Mantel test, see also Figure S5). Then, we choose to quantify genetic friction using the F_{ST} between a population living exactly in the barycentric center of the county and a putative neighboring population living 30 km away (see Figure S6). We consider the pairwise $(1 - F_{ST})$ values between the counties as input matrix of pairwise similarities (Σ_{xx}). We find that the northernmost counties (Nordbotten, Västerbotten and Jämtlands) have the strongest friction values whereas the smallest values were found in the regions around the Stockholm area (Östergötlands, Stockholms, Södermanlands, Västmanlands and Jönköpings are the five counties with the lowest friction values). The fourth largest value is in the Dalarna county (in north middle Sweden), which borders southern Norway, and counts more individuals with remote Finnish or Norwegian ancestry than other counties (HUMPHREYS *et al.* 2011). As expected, the four counties with the largest friction values are also the most differentiated from other counties even when controlling for geographic distance (Figure S5).

In summary, we confirm the results of HUMPHREYS *et al.* (2011) who found that there is more genetic differentiation within northern Sweden than within southern Sweden. The four counties with the largest friction values are also the counties with the lowest population densities (Figure S7) suggesting that low density habitat triggered population differentiation in Northern Sweden.

Non-stationary patterns of IBD for 20 alpine plants: We consider a set of 20 alpine plant species that have been sampled across the Alps (GUGERLI *et al.* 2008; ALVAREZ *et al.* 2009; JAY *et al.* 2012). The sampling is particularly dense with three individuals per species collected and genotyped for each cell of approximately 500 km². The individual genotype consists of amplified fragment length polymorphisms (AFLPs). Because of the individual-based sampling, we consider the matrix of correlation between the individual genotypes as input data matrix (Σ_{xx}). The genetic friction that we estimate corresponds to the expected correlation between sampled individuals and neighboring individuals located at 10 km.

The map of genetic friction for *Rhododendron ferrugineum*, one of the 20 sampled plant species, exhibits large friction in the southern part of a corridor that cuts the main chain of the Alps and roughly stretches from Innsbruck to Lake Garda (Figure 7). This pattern is representative of what is generally found for the other species because this region corresponds to the main zone of genetic friction for 19 out of the 20 alpine species we consider (see Figure S8 for all friction maps). When using a log scale for the friction values, we find for most of the species that the whole corridor, and not its Southern part only, has more elevated friction values than the rest of the Alps (Figure S9). The southern part of the corridor corresponds to the Adige valley, which is perpendicular to the main E-W axis of the Alps and which further leads to the Reschen and the Brenner pass, the later being the lower pass across the main chain of the Alps (TAYLOR 1940). For plant species, the Adige valley—also referred as the *Brenner line*— is already known to be an important barrier for plants both at the intraspecific and interspecific level (THIEL-EGENTER *et al.* 2011). The explanation for the presence of a barrier in the Brenner line involves Pleistocene glaciations: the populations of plants were initially fragmented into glacial refuges, then expanded via postglacial colonization routes, and a secondary contact zone finally arose around the Brenner line where formerly allopatric populations admixed (PAWLOWSKI 1970; SCHÖNSWETTER *et al.* 2005; THIEL-EGENTER *et al.* 2011).

To check that the observed pattern is not an artifact of the sampling scheme, we simulate a pairwise matrix of correlation values (Σ_{xx}) using the stationary model of equation (4) which assumes that correlation decays at the same rate in the entire habitat. We consider values for the parameters of the correlogram that provide a good fit to the observed decay of correlation in *Rhododendron ferrugineum*. The inferred map of genetic friction indicates that the kriging method correctly infers homogeneous friction, which confirms that the large friction found around the Adige river is not an artifact of the kriging method (Figure 7).

DISCUSSION

In this article we present a new Bayesian method to characterize non-stationary patterns of isolation-by-distance. From *global* measures of pairwise similarity or dissimilarity, the method infers *local* measures of similarity or dissimilarity. Whatever is the exact measure of genetic (dis)similarity, we use the generic expression of *genetic friction* when referring to the estimated local growth of genetic dissimilarity or differentiation. If considering for instance the F_{ST} pairwise matrix of genetic differentiation between populations, the inferred genetic friction correspond to the F_{ST} between the sampled populations and fictive neighboring populations located at a given distance. The method is not restricted to F_{ST} measures and can handle any type of measures of differentiation and is also valid at the individual scale. We consider for instance the correlation between the allelic types of individuals, but other measures would be valid such as identity by descent between individuals (BROWNING and BROWNING 2011) as well as coancestry measures (LAWSON *et al.* 2012). Because the two latter measures are based on haplotypes instead of genotypes, they can

provide information at a finer scale (GATTEPAILLE and JAKOBSSON 2012; LAWSON *et al.* 2012).

Genetic differentiation and gene flow: It is of course tempting to convert maps of genetic friction into maps of gene flow or of dispersal distance. Assuming that differentiation occurs according to a stepping stone model, such parameter estimates could be obtained using theoretical relationships between local F_{ST} and dispersal distance (ROUSSET 1997). However, it is well known that relating F_{ST} or other measures of genetic differentiation to gene flow relies on many assumptions that may be unrealistic (MARKO and HART 2011). Although the estimation of gene flow with F_{st} -based methods can be robust in some situations, such as temporal variation of gene flow (LEBLOIS *et al.* 2004), there are other processes such as range expansion, local extinction and recolonization that can modify drastically the pattern of genetic differentiation (WADE and MCCAULEY 1988; EXCOFFIER *et al.* 2009; ARENAS *et al.* 2012; SLATKIN and EXCOFFIER 2012). More generally, providing a map of genetic friction is informative about the *pattern* of genetic differentiation but does not distinguish among the evolutionary *processes* that generated this pattern (for similar concerns about PCA, see MCVLEAN 2009). To provide a concrete example, we find that the same pattern, a zone of elevated genetic friction, can correspond to a barrier to gene flow (Figures 2-4) or to a secondary contact zone in the case of the alpine species (Figure 7).

Relevance to landscape genetics: The two key steps of landscape genetics are the detection of genetic discontinuities and the correlation of these discontinuities with landscape and environmental features such as barriers (MANEL *et al.* 2003). Detection of genetic discontinuities is clearly provided by the Kriging method; for instance in the case of the alpine species, we find genetic discontinuities, i.e. larger genetic friction, around an important Alpine valley (Figure 7). The second key step where these discontinuities are correlated with landscape or environmental variable can also be obtained as a post-processing step by correlating estimated genetic friction with landscape variables. For instance, in the case of the human SNP Swedish data, we find that genetic friction is correlated with population density. There are alternative and *integrative* approaches that account for both genetic data and landscape variables within the same statistical framework. Accounting for both sources of data can be performed either by a joint assessment of the pattern of population structure or differentiation and its correlation with landscape or environmental variable (FOLL and GAGGIOTTI 2006; JAY *et al.* 2011), or by correlating genetic distances with distances based on landscape features (CUSHMAN *et al.* 2006; MCRAE 2006). These integrative approaches are *hypothesis-driven* in the sense that each set of landscape features affecting population structure corresponds to one hypothesis that can be tested or compared to other ones. The proposed kriging approach is instead a technique of *exploratory data analysis*. It might be especially appropriate for large-scale conservation studies not focused on the underlying evolutionary processes but that should deal with reserve design and with the management of fragmented populations (SCHWARTZ *et al.* 2007). To compare or test the support of different evolutionary processes provided by the pattern of non-stationary IBD, we should rather resort to inference based on explicit simulations of evolutionary processes using for instance approximate Bayesian computation (CSILLÉRY *et al.* 2010). Within this simulation framework, measures of genetic friction can be included

as statistical summaries of the data.

Caveats: As with other descriptive techniques of exploratory data analysis, there is no hypothesis-testing or goodness of fit procedures attached to it. This is a concern of primary importance because it can lead to overinterpretations of friction maps. When analyzing the alpine plant species data, we proposed one solution. We simulated spatially homogeneous patterns of genetic differentiation using the same sampling scheme as in the data and we checked that the obtained friction pattern is indeed more homogeneous than the one obtained with the data (Figure 7). Using the same sampling data as in the data is crucial since the sampling scheme affects the ascertainment of population structure for various statistical methods (MCVEAN 2009; SCHWARTZ and MCKELVEY 2009). An additional concern about the kriging method concerns the choice of the distance between the sampled populations or individuals and the fictive neighboring populations or individuals. We considered various choices of distances for both the Swedish data set (10-150 km) and one of the alpine species (5-100km) and found that these large ranges of distances provided similar patterns of genetic friction although the patterns generated with the larger distances were smoother (Figures S10 and S11). The last caveat to bear in mind concerns the choice of the pairwise dissimilarity or differentiation matrix. For instance, in a simulation study with 5 populations, LAWSON and FALUSH (2012) showed that the five populations were clearly distinguishable with some but not all pairwise dissimilarity matrices between individuals. Although a potential caveat, being able to choose the measure of dissimilarity also adds to the flexibility of the method and different measures can convey information about processes that occurred at different time periods.

Perspectives: When carefully addressing the aforementioned caveats, measures of genetic friction can provide interpretable patterns for describing non-stationary patterns of IBD. The method relies on a matrix of pairwise (dis)similarity between individuals or populations located on georeferenced sampling sites. An approach based on dissimilarity matrices is an appropriate methodology for the new genomic era where we have to deal with massive data. Computing pairwise dissimilarity matrix can be computationally efficient (e.g. BROWNING and BROWNING 2011) and can even be parallelized to compute different parts of the matrix (LAWSON and FALUSH 2012). Having a statistical method able to scale with the dimension of the genetic data should make it a valuable tool for investigating patterns of genetic differentiation in a wide range of studies.

LITERATURE CITED

- ALVAREZ, N., C. THIEL-EGENTER, A. TRIBSCH, R. HOLDEREGGER, S. MANEL, P. SCHÖNSWETTER, P. TABERLET, S. BRODBECK, M. GAUDEUL, L. GIELLY, *et al.*, 2009 History or ecology? substrate type as a major driver of spatial genetic structure in Alpine plants. *Ecology Letters* **12**: 632–640.
- ARENAS, M., N. RAY, M. CURRAT, and L. EXCOFFIER, 2012 Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution* **29**: 207–218.

- ARNAUD, J., 2003 Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa*: influence of landscape heterogeneity. *Landscape Ecology* **18**: 333–346.
- BARBUJANI, G., N. ODEN, and R. SOKAL, 1989 Detecting regions of abrupt change in maps of biological variables. *Systematic Biology* **38**: 376–389.
- BISHOP, C., 2006 *Pattern recognition and machine learning*, volume 4. springer New York.
- BOCQUET-APPEL, J. and J. BACRO, 1994 Generalized wombling. *Systematic Biology* **43**: 442–448.
- BROWNING, B. and S. BROWNING, 2011 A fast, powerful method for detecting identity by descent. *The American Journal of Human Genetics* **88**: 173–182.
- CASTELLA, V., M. RUEDI, L. EXCOFFIER, C. IBANEZ, R. ARLETTAZ, and J. HAUSSE, 2000 Is the Gibraltar Strait a barrier to gene flow for the bat *Myotis myotis* (Chiroptera: Vespertilionidae)? *Molecular Ecology* **9**: 1761–1772.
- CERCUEIL, A., O. FRANÇOIS, and S. MANEL, 2007 The genetical bandwidth mapping: a spatial and graphical representation of population genetic structure based on the Wombling method. *Theoretical population biology* **71**: 332–341.
- CRESSIE, N. A. C., 1993 *Statistics for Spatial Data (Wiley Series in Probability and Statistics)*. Wiley-Interscience.
- CRIDA, A. and S. MANEL, 2007 Wombsoft: an r package that implements the wombling method to identify genetic boundary. *Molecular Ecology Notes* **7**: 588–591.
- CSILLÉRY, K., M. G. B. BLUM, O. E. GAGGIOTTI, and O. FRANÇOIS, 2010 Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution* **25**: 410–418.
- CUSHMAN, S., K. MCKELVEY, J. HAYDEN, and M. SCHWARTZ, 2006 Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist* **168**: 486–499.
- DUPANLOUP, I., S. SCHNEIDER, and L. EXCOFFIER, 2002 A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**: 2571–2581.
- EPPS, C., P. PALSBOELL, J. WEHAUSEN, G. RODERICK, R. RAMEY II, and D. MCCULLOUGH, 2005 Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology Letters* **8**: 1029–1038.
- EXCOFFIER, L., M. FOLL, and R. PETIT, 2009 Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics* **40**: 481–501.

- FOLL, M. and O. GAGGIOTTI, 2006 Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**: 875–891.
- GATTEPAILLE, L. M. and M. JAKOBSSON, 2012 Combining markers into haplotypes can improve population structure inference. *Genetics* **190**: 159–174.
- GAUFFRE, B., A. ESTOUP, V. BRETAGNOLLE, and J.-F. COSSON, 2008 Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology* **17**: 4619–4629.
- GUGERLI, F., T. ENGLISCH, H. NIKLFELD, A. TRIBSCH, Z. MIREK, M. RONIQUIER, N. ZIMMERMANN, R. HOLDEREGGER, and P. TABERLET, 2008 Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation—a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics* **10**: 259–281.
- HANDCOCK, M. and M. STEIN, 1993 A Bayesian analysis of kriging. *Technometrics* pp. 403–410.
- HARDY, O. J., L. MAGGIA, E. BANDO, P. BREYNE, H. CARON, M. H. CHEVALLIER, A. DOLIGEZ, C. DUTECH, A. KREMER, C. LATOUCHE-HALLÉ, *et al.*, 2006 Fine-scale genetic structure and gene dispersal inferences in 10 neotropical tree species. *Molecular Ecology* **15**: 559–571.
- HELLBERG, M., 2009 Gene flow and isolation among populations of marine animals. *Annu. Rev. Ecol. Evol. Syst.* **40**: 291–310.
- HUDSON, R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.
- HUMPHREYS, K., A. GRANKVIST, M. LEU, P. HALL, J. LIU, S. RIPATTI, K. REHNSTRÖM, L. GROOP, L. KLARESKOG, B. DING, *et al.*, 2011 The genetic structure of the Swedish population. *PloS one* **6**: e22547.
- JAY, F., O. FRANÇOIS, and M. G. B. BLUM, 2011 Predictions of native American population structure using linguistic covariates in a hidden regression framework. *PloS one* **6**: e16227.
- JAY, F., S. MANEL, N. ALVAREZ, E. Y. DURAND, W. THUILLER, R. HOLDEREGGER, P. TABERLET, and O. FRANÇOIS, 2012 Forecasting changes in population genetic structure of alpine plants in response to global warming. *Molecular Ecology* .
- LAWSON, D., G. HELLENTHAL, S. MYERS, and D. FALUSH, 2012 Inference of population structure using dense haplotype data. *PLoS Genetics* **8**: e1002453.
- LAWSON, D. J. and D. FALUSH, 2012 Population identification using genetic data. *Annual review of genomics and human genetics* **Advance access**.

- LE, N. D. and J. V. ZIDEK, 1992 Interpolation with uncertain spatial covariances: A Bayesian alternative to kriging. *Journal of Multivariate Analysis* **43**: 351–374.
- LEBLOIS, R., F. ROUSSET, and A. ESTOUP, 2004 Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics* **166**: 1081–1092.
- MANEL, S., F. BERTHOUD, E. BELLEMAIN, M. GAUDEUL, G. LUIKART, J. SWENSON, L. WAITS, P. TABERLET, *et al.*, 2007 A new individual-based spatial approach for identifying genetic discontinuities in natural populations. *Molecular Ecology* **16**: 2031–2043.
- MANEL, S., M. SCHWARTZ, G. LUIKART, and P. TABERLET, 2003 Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**: 189–197.
- MANNI, F., E. GUERARD, E. HEYER, *et al.*, 2004 Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier’s algorithm. *Human Biology* **76**: 173–190.
- MARKO, P. and M. HART, 2011 The complex analytical landscape of gene flow inference. *Trends in ecology & evolution* .
- MCRAE, B., 2006 Isolation by resistance. *Evolution* **60**: 1551–1561.
- MCRAE, B. and P. BEIER, 2007 Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences* **104**: 19885.
- MCVEAN, G., 2009 A genealogical interpretation of principal components analysis. *PLoS Genetics* **5**: e1000686.
- MUNSHI-SOUTH, J., 2012 Urban landscape genetics: canopy cover predicts gene flow between white-footed mouse (*Peromyscus leucopus*) populations in New York City. *Molecular Ecology* .
- PACIOREK, C. J. and M. J. SCHERVISH, 2006 Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17**: 483–506.
- PAWLOWSKI, B., 1970 Remarques sur l’endémisme dans la flore des alpes et des carpates. *Plant Ecology* **21**: 181–243.
- PICKETT, S. and M. CADENASSO, 1995 Landscape ecology: spatial heterogeneity in ecological systems. *Science* **269**: 331–334.
- RAMACHANDRAN, S., O. DESHPANDE, C. C. ROSEMAN, N. A. ROSENBERG, M. W. FELDMAN, and L. L. CAVALLI-SFORZA, 2005 Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 15942–15947.

- RAY, N., 2005 PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Molecular Ecology Notes* **5**: 177–180.
- RILEY, S., J. POLLINGER, R. SAUVAJOT, E. YORK, C. BROMLEY, T. FULLER, and R. WAYNE, 2006 Fast-track: A southern california freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology* **15**: 1733–1741.
- ROUSSET, F., 1997 Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- SAFNER, T., M. MILLER, B. MCRAE, M. FORTIN, and S. MANEL, 2011 Comparison of bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences* **12**: 865–889.
- SCHMIDT, A. and A. O’HAGAN, 2003 Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**: 743–758.
- SCHÖNSWETTER, P., I. STEHLIK, R. HOLDEREGGER, and A. TRIBSCH, 2005 Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology* **14**: 3547–3555.
- SCHWARTZ, M., G. LUIKART, and R. WAPLES, 2007 Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution* **22**: 25–33.
- SCHWARTZ, M. and K. MCKELVEY, 2009 Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics* **10**: 441–452.
- SERROUYA, R., D. PAETKAU, B. N. MCLELLAN, S. BOUTIN, M. CAMPBELL, and D. A. JENKINS, 2012 Population size and major valleys explain microsatellite variation better than taxonomic units for caribou in western Canada. *Molecular Ecology* **21**: 2588–2601.
- SHARBEL, T., B. HAUBOLD, and T. MITCHELL-OLDS, 2000 Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology* **9**: 2109–2118.
- SLATKIN, M., 1985 Gene flow in natural populations. *Annual review of ecology and systematics* **16**: 393–430.
- SLATKIN, M., 1993 Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264–279.
- SLATKIN, M. and L. EXCOFFIER, 2012 Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics* **23**: 171–181.

- STORFER, A., M. MURPHY, S. SPEAR, R. HOLDEREGGER, and L. WAITS, 2010 Landscape genetics: where are we now? *Molecular Ecology* **19**: 3496–3514.
- TAYLOR, G., 1940 Trento to the Reschen pass: a cultural traverse of the Adige corridor. *Geographical Review* **30**: 215–237.
- THIEL-EGENTER, C., N. ALVAREZ, R. HOLDEREGGER, A. TRIBSCH, T. ENGLISCH, T. WOHLGEMUTH, L. COLLI, M. GAUDEUL, L. GIELLY, N. JOGAN, *et al.*, 2011 Break zones in the distributions of alleles and species in Alpine plants. *Journal of Biogeography* **38**: 772–782.
- WADE, M. and D. MCCAULEY, 1988 Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* pp. 995–1005.
- WRIGHT, S., 1943 Isolation by distance. *Genetics* **28**: 114–138.
- ZALEWSKI, A., S. PIERTNEY, H. ZALEWSKA, and X. LAMBIN, 2009 Landscape barriers reduce gene flow in an invasive carnivore: geographical and local genetic structure of American mink in Scotland. *Molecular Ecology* **18**: 1601–1615.

APPENDIX A: COMPUTING THE CORRELATION BETWEEN SAMPLED AND UNSAMPLED SITES

To renormalize equation (3), we have to compute the diagonal elements of the covariance matrix Σ_{yy} for the unsampled sites. Using the fact that the residuals of the regression equation (2) are of variance $\Psi_{yy} - \Psi_{xy}\Psi_{xx}^{-1}\Psi_{xy}$ (BISHOP 2006), we can show that the covariance matrix Σ_{yy} is given by

$$\Sigma_{yy} = \tau \Sigma_{xx} \tau^T + \Psi_{yy} - \Psi_{xy} \Psi_{xx}^{-1} \Psi_{xy}.$$

APPENDIX B: GIBBS SAMPLER

For the parameters of the correlogram model, we choose the following prior distributions

$$\alpha \sim \mathcal{U}(\min S_{ij} - \delta, \min S_{ij} + \delta) \quad (5)$$

$$\lambda \sim \log(\mathcal{U}(0, 1)) \quad (6)$$

$$r \sim \mathcal{U}(\min D_{ij}, \max D_{ij}), \quad (7)$$

where S is the empirical covariance matrix and D denotes the pairwise geographical distance between sampled sites. Here, we adopt an empirical Bayes approach, since we partly use the data when defining the prior distributions.

Using the Bayes formula we have

$$\log(P(\alpha, \lambda, r|S)) = \log(p(S|\alpha, \lambda, r)) + \log(p(\alpha, \lambda, r)) + C_1, \quad (8)$$

where C_1 is a constant and the likelihood is given by the Wishart distribution (SCHMIDT and O'HAGAN 2003)

$$\log(p(S|\alpha, \lambda, r)) = -\frac{l}{2} \log(|\Psi_{xx}|) - \frac{1}{2} \text{trace}(\Psi_{xx}^{-1} S) + C_2, \quad (9)$$

where C_2 is an other constant. We then simulate a sample of the joint posterior probability using the Gibbs sampling algorithm. To obtain replicates from each conditional distribution, we consider a grid for each parameter. We evaluate the conditional densities for each point of the grid using equations (8) and (9) and then we use a sampling algorithm for simulating discrete random variables with known probability masses.

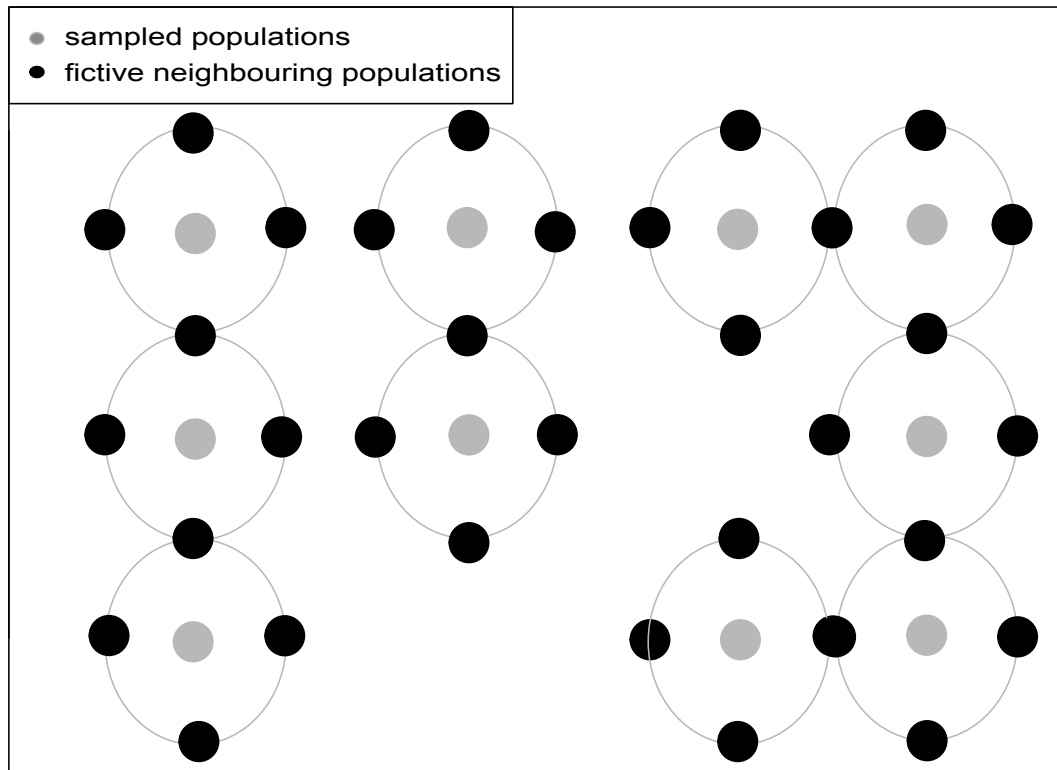


Figure 1: A 2-dimensional habitat with putative sampled locations and neighboring locations in grey. Local differentiation at each sampled location corresponds to the average (over neighbors) pairwise measure of differentiation between the population or individual at the sampled location and its neighbors.

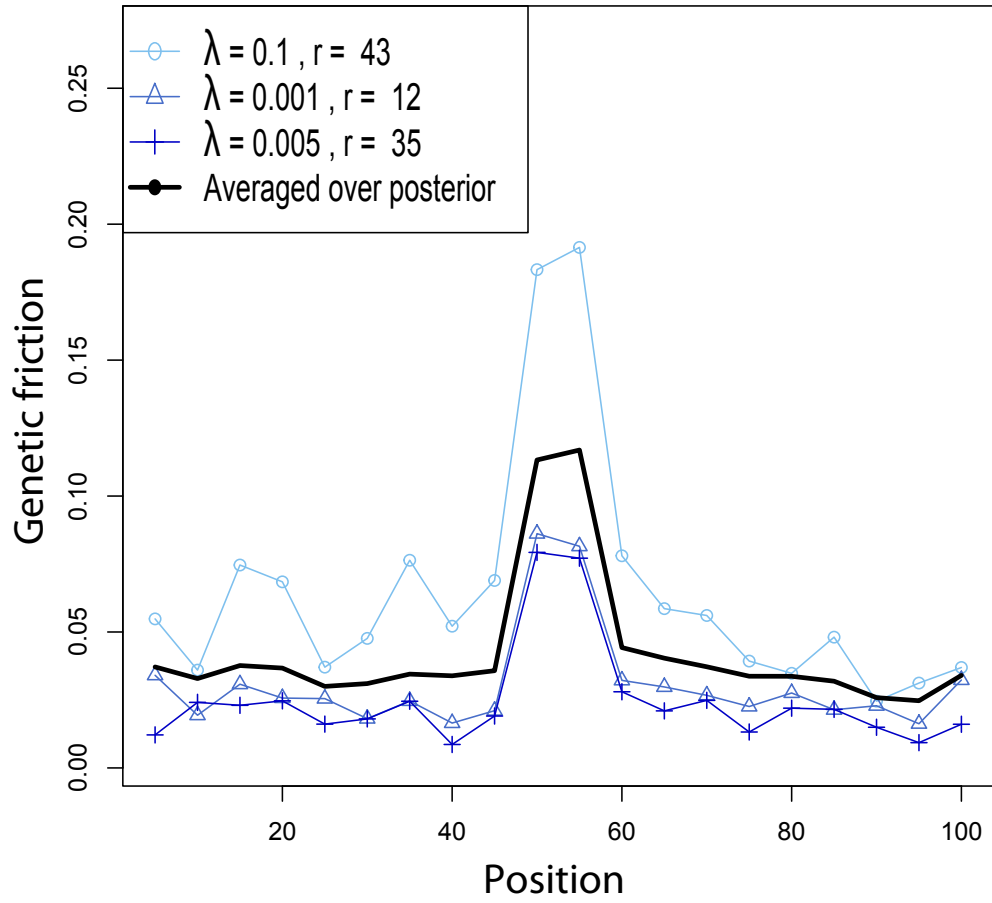


Figure 2: Estimation of the friction in a one-dimensional habitat with a genetic barrier in the middle of the habitat. Genetic friction corresponds to one minus the correlation of allele frequencies between sampled and unsampled neighboring demes. The genetic barrier is located between the deme 50 and 51.

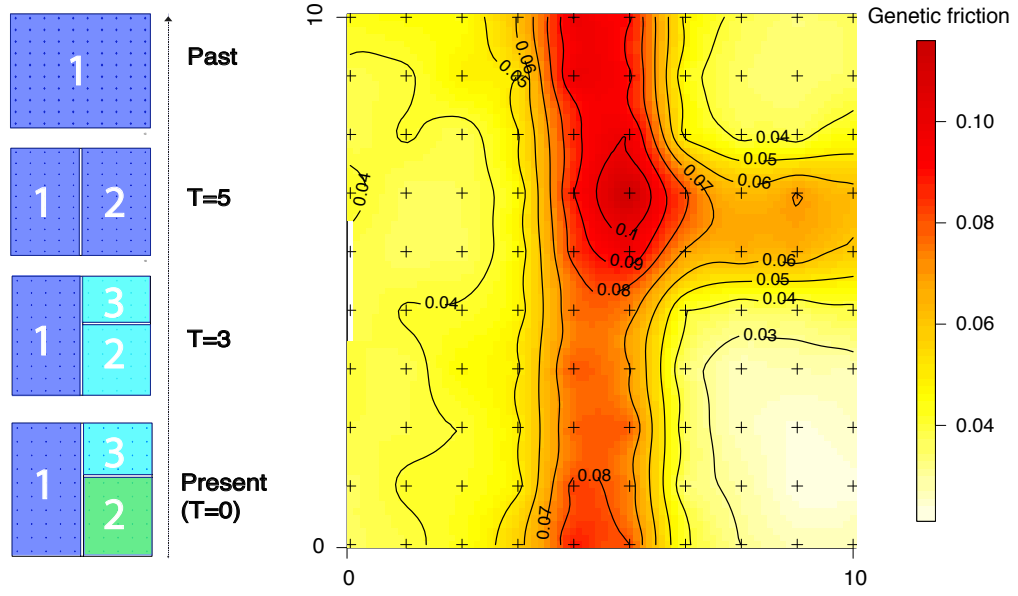


Figure 3: Friction map in a two-dimensional habitat with two genetic barriers. On the left hand side, the time line shows the time when the genetic barriers appeared.

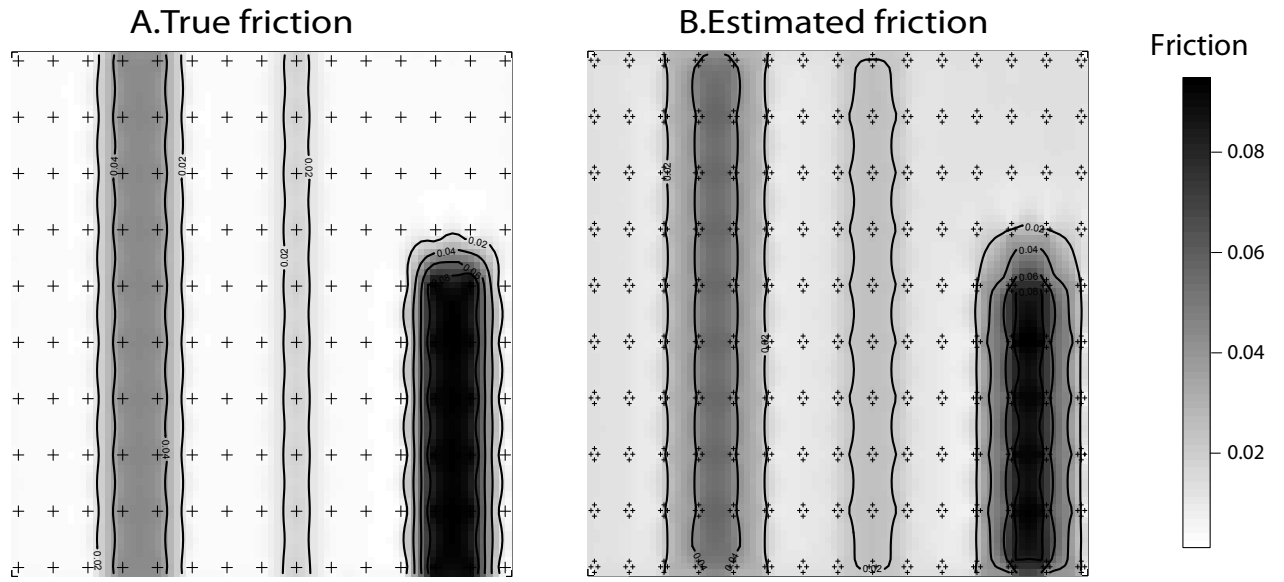


Figure 4: True and estimated friction maps in a two dimensional habitat of dimension 15×10 containing three barriers. The friction at sampled points corresponds to one minus the correlation between sampled points (large crosses in panel A) and fictive neighbors (small crosses in panel B) located at a distance of 0.1. The correlation landscape is specified with the non-stationary model of PACIOREK and SCHERVISH (2006) considering smaller range values (parameter r of equation 4) for the zones of abrupt genetic change. The true friction of panel A is obtained using the convolution formula of PACIOREK and SCHERVISH (2006) at distance 0.1 which provides pairwise correlation based on the landscape of the range parameter of equation (4).

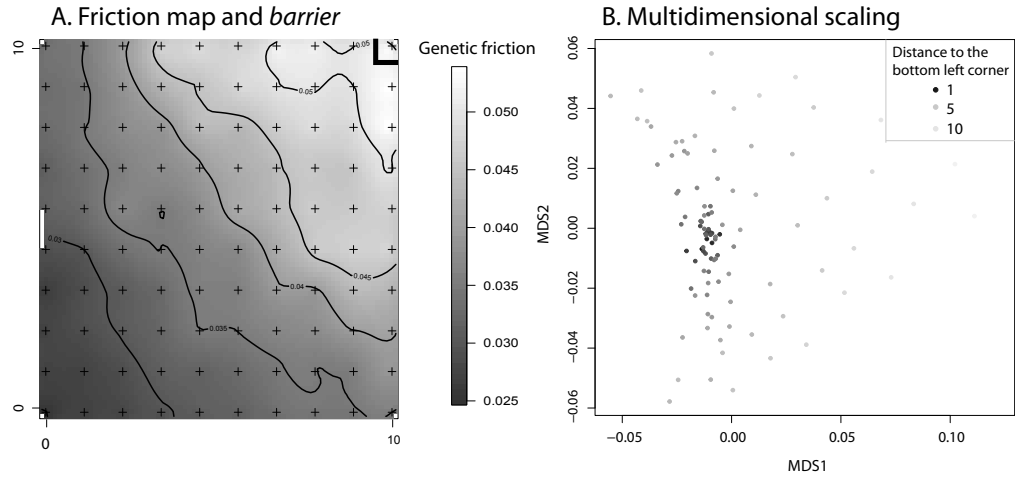


Figure 5: Investigating the pattern of population differentiation for a gradient of gene flow in a two dimensional habitat. Gene flow is maximal at the lower left corner of the habitat and decreases proportionally to the distance from the lower left corner. Panel A) Estimation of friction values using the correlation with neighbors located at a distance of 0.1. The first genetic barrier that is found with the software *barrier* is shown with a thick black line. Panel B) Multidimensional scaling plot with a grayscale color scheme to represent the distance of each population to the lower left corner. Dark points correspond to populations that are close to the lower left corner whereas the lightest points are the farthest away from the lower left corner.

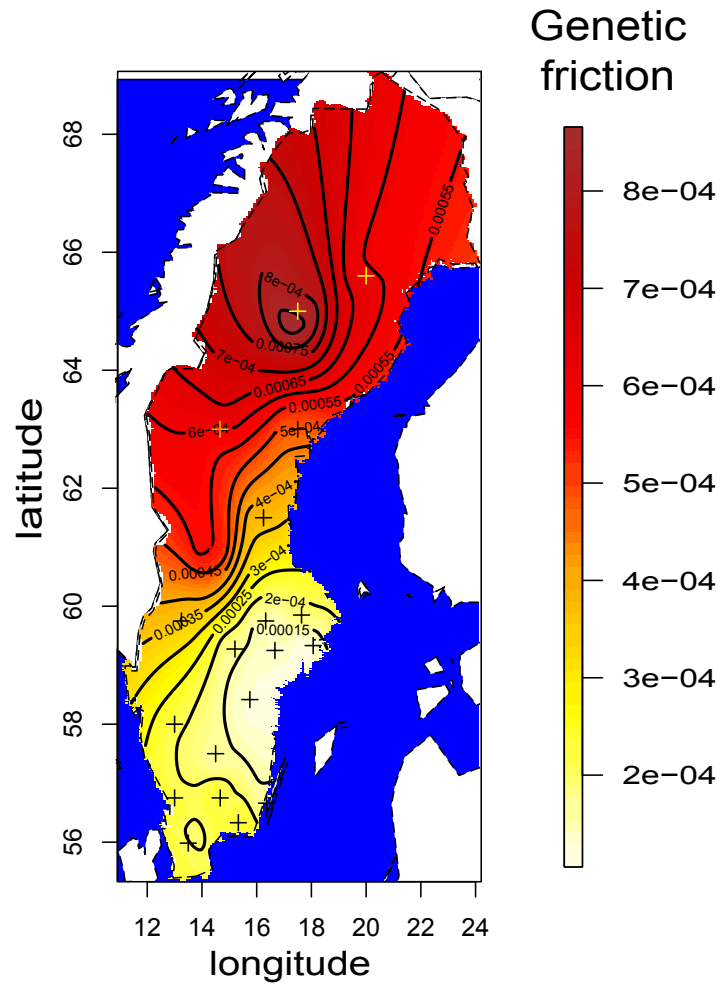


Figure 6: Map of genetic friction computed for the 20 Swedish counties. Genetic friction corresponds to the F_{st} between sampled county and fictive neighboring populations, not shown, located at 30 km.

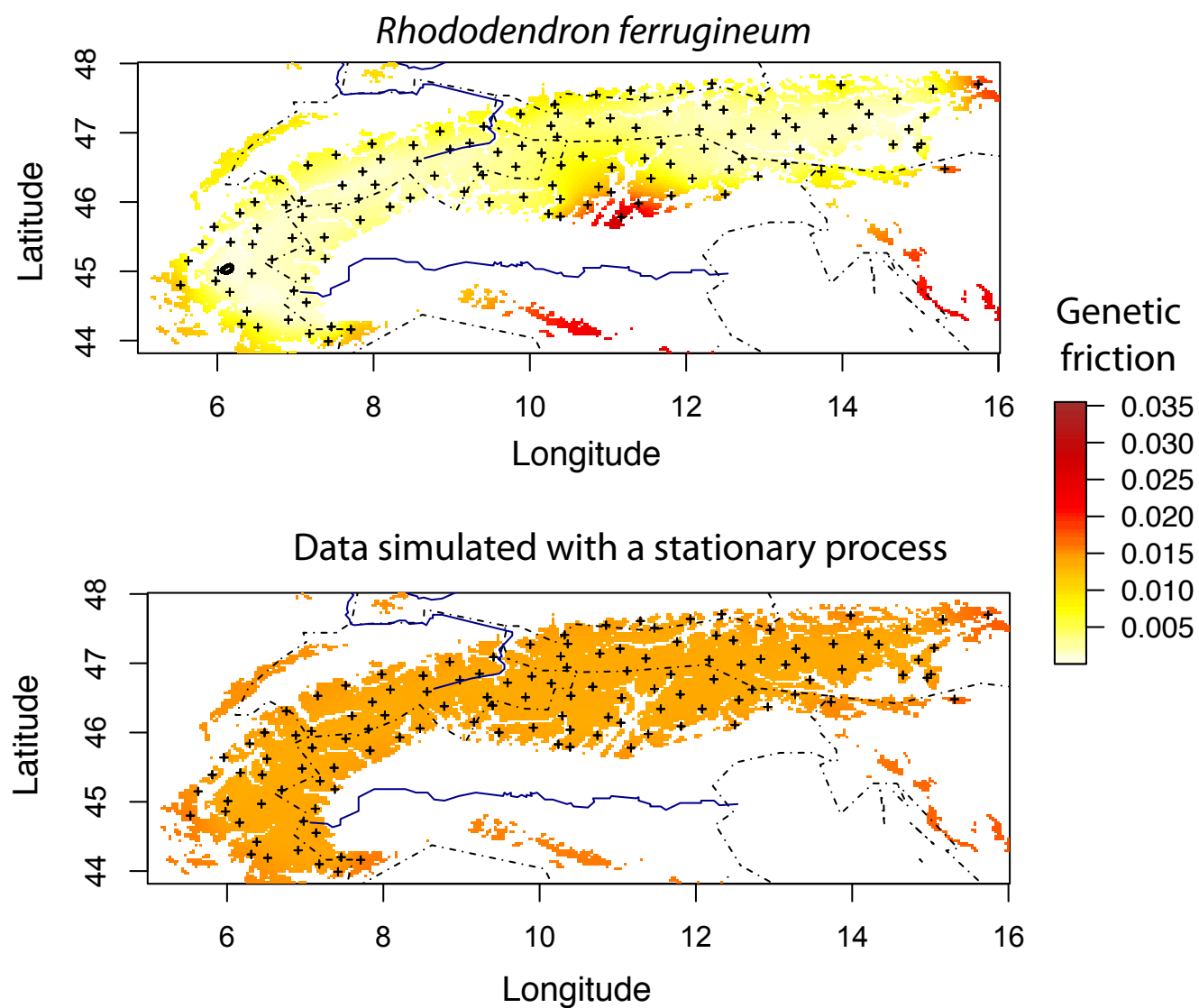


Figure 7: Map of genetic friction computed across the Alps for *Rhododendron ferrugineum*. Genetic frictions corresponds to the correlation between sampled individuals and fictive neighboring individuals, not shown, located at 10 km around the sampled individuals.